

Robust Head Pose Estimation Using Contourlet Transform

Mohammad Tofighi, Hashem Kalbkhani, Mahrokh G. Shayesteh, and Mehdi Ghasemzadeh

Department of Electrical Engineering, Urmia University, Urmia, Iran

Emails: mo.tofighi@gmail.com, {st_h.kalbkhani, m.shayesteh}@urmia.ac.ir, me_ghasemzadeh@yahoo.com

Abstract—Head pose estimation is an important pre-processing step in many pattern recognition and computer vision systems such as face recognition. Since the performance of the face recognition systems is greatly affected by the pose of the face, how to estimate the accurate pose of the face in the face image is still a challenging problem. In this paper, we present a novel method for head pose estimation. To enhance the efficiency of the estimation we first use contourlet transform for feature extraction. Contourlet transform is a multi-resolution, multi-direction transform. Finally, in order to reduce the feature space dimension and obtain appropriate features, we use LDA (Linear Discriminant Analysis) and PCA (Principal Component Analysis) to remove inefficient features. Then, we apply k-nearest neighborhood (k-NN) and minimum distance classifiers to classify the pose of head. We use the public available FERET database to evaluate the performance of the proposed method. Simulation results indicate the efficiency of the proposed method in comparison with previous method.

Index Terms—Contourlet transform, Head pose estimation, k-NN, LDA, PCA.

I. INTRODUCTION

Head pose estimation is the process of extracting the orientation of a face or head from an image containing the face. Pose estimation is always used by human on relations to show a heightened sense of awareness and understanding by non-verbal communications when interacting with other people. There has been a significant improvement in face recognition over the last two decades. However, robust and accurate face recognition is still a classic problem because of variations in the pose of the face. Therefore, pose invariant face perception has been an active research topic for several years.

Several methods were introduced in the literature in the late 1990's and it continues up to now [1]. The proposed methods had advantages and also many disadvantages, and to overcome the disadvantages and enhance the advantages, many studies have been done in this field. The introduced methods can be classified into four categories as: 1) Template matching [2] methods which use the nearest neighbor classification to find the most similar view of a new head pose, 2) Appearance-based methods [3] that apply pattern classification or nonlinear regression tools, develop a functional mapping from the image or feature data to a head pose measurement, 3) Geometric methods [4] that determine poses by using the relative configuration of facial landmarks (such as eyes, mouth, nose tip), 4) Dimensionality reduction methods [5]–[7] which seek a low-dimensional continuous manifold constrained by the pose

variations, and then new images can be embedded into these manifolds and used for template matching or regression.

In this paper, we develop a new method to estimate the pose of the head in the face image. In our proposed method, we use contourlet transform to extract the features from the image. Moreover, to reduce the dimension of the feature space and also, to increase the separation we utilize Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). In the final step, we apply the extracted features to the classifiers such as k-nearest neighborhood (k-NN) and minimum distance to classify them.

The rest of the paper is organized as follows: Section II introduces the contourlet transform. Section III explains the proposed method. Simulation results and comparisons with other pose estimator systems are presented in section IV. Finally, section V concludes the paper.

II. CONTOURLET TRANSFORM

Minh *et al.* [8] developed Contourlet transform based on an efficient two-dimensional (2-D) multiscale and directional filter bank that can deal effectively with images having smooth contours. Contourlets not only possess the main features of wavelets (namely multiscale and time-frequency localization), but also offer a high degree of directionality and anisotropy. The main difference between the contourlets and the other multiscale directional systems is that the contourlet transform allows for different and flexible number of directions at each scale, while achieving nearly critical sampling rate.

In contourlet transform to achieve a sparse series for images with smooth contours and having useful computations, the iterated filter banks are utilized. Contourlet transform is composed of Laplacian Pyramid (LP) [9], [10] and Directional Filter Bank (DFB) [11]. The first stage in this transform is to capture the points discontinuities by Laplacian Pyramid, and then followed by a directional filter bank to relate point discontinuities into linear structures. At each level, the LP decomposition generates a downsampled lowpass version of the original image, and then a bandpass image is resulted from the difference between the original image and the prediction one. Subsequently, the bandpass signal is filtered in different directions. The singular points in the same direction are synthesized to a coefficient by DFB.

As shown in Fig. 1(a), DFB divides the highpass image into directional subbands in the frequency domain. Fig. 1(a) shows

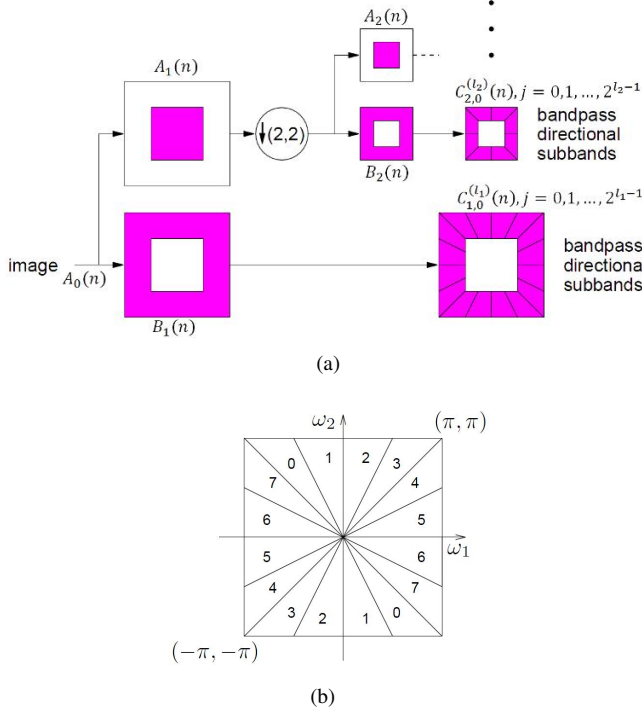


Fig. 1. Spectrum partitioning, (a) decomposition scheme of the contourlet transform, (b) 8-subband partition.

a multiresolution and directional decomposition of contourlet filter bank which is composed of an LP and a DFB. To capture the directional information, bandpass images from the LP are fed into a DFB. Fig. 1(b) shows the DFB with level $l = 3$ where there are $2^3 = 8$ real wedge-shaped frequency bands. Subbands 0-3 correspond to the mostly horizontal directions, while subbands 4-7 correspond to the mostly vertical directions.

The DFB divides the highpass image into 2^l wedge-shaped frequency bands after l level tree-structure segmentation. Since the DFB was designed to capture the high frequency of the input image, the low frequency content is poorly handled. In fact, the DFB alone does not provide a sparse representation for images, because with the frequency partition shown in Fig. 1(b), low frequency would leak into several directional subbands. This fact provides another reason to combine the DFB with a multiresolution decomposition, where low frequencies of the input image are removed before applying the DFB.

Let $A_0(n)$ be the input image as shown in Fig. 1(a). The output after the LP decomposition is a lowpass image $A_1(n)$ and I bandpass images $B_i(n)$, $i = 1, 2, \dots, I$. That means, the i -th level of LP decomposes the image $A_{i-1}(n)$ into a coarser image $A_i(n)$ and a highpass image $B_i(n)$. Then each of highpass images $B_i(n)$ are further decomposed by a l_i -level DFB into 2^{l_i} highpass directional images $C_{y,j}^{(l_i)}(n)$, $j = 0, 1, \dots, 2^{l_i-1}$. The combined result is a double iterated filter bank structure, named contourlet filter bank, that is resulted in contourlet transform.

Fig. 2 presents the diagram that shows the sequence we used in our proposed head pose estimation algorithm. In order to compute contourlet transform of head image, we use the contourlet transform toolbox provided in [12]. In our method, at first, color face image is converted to gray-scale image. To do this, RGB image is converted to gray-scale by averaging the r , g , and b components of the color image. This method eliminates the hue and saturation information while retaining the luminance. We use pyramidal directional filter bank (PDFB) decomposition; ‘9-7’ filters for the multiscale decomposition stage and ‘PKVA’ filters for the directional decomposition stage [13], [14]. There will be two levels of pyramidal decomposition and the numbers of directional decomposition at each pyramidal level are set to 0 and 1. Using this decomposition, we obtain six sub-bands. These sub-bands represent our feature space. Fig. 4 shows a typical image and the sub-bands obtained from the contourlet transform.

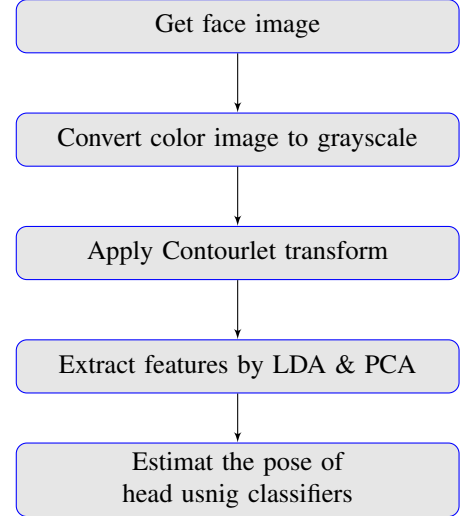


Fig. 2. Block diagram of the proposed head pose estimation algorithm

All the extracted features are not appropriate for our work; consequently, the improper features should be removed from the feature space. There are several approaches for feature space dimension reduction, such as PCA and LDA. If we use only LDA, sometimes scatter matrix in LDA becomes singular and this degrades the performance of system. Hence, we use PCA before LDA to overcome this problem. PCA is unsupervised approach which is used only for whitening. LDA is supervised algorithm, thus it is good approach for feature space dimension reduction. More details are presented in the experimental results section.

After obtaining the best feature vector, the final step is classifying the features. In this work, we choose two different classifiers: minimum distance and k-nearest neighborhood (k-NN). More details are included in the next section.

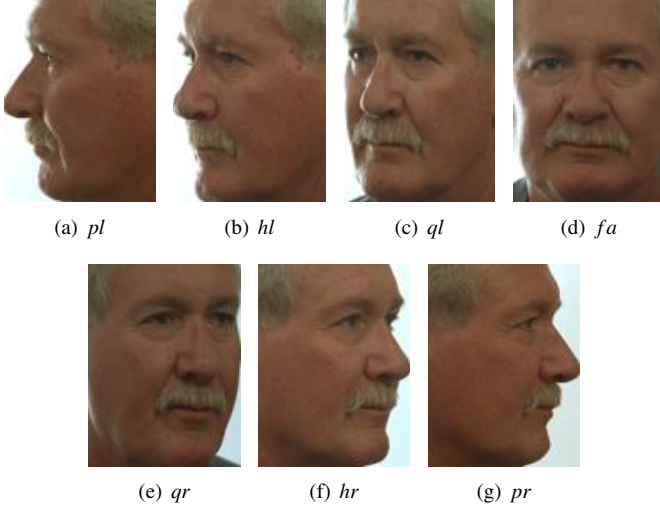


Fig. 3. Different head poses with label in FERET database

IV. SIMULATION RESULTS AND DISCUSSION

A. Database

We need the database that contains face images with different rotations. So, we choose FERET database [15]. In this database, lighting conditions, face size and descent of people vary too much. The images have different indices. From different indices, we select face rotations represented by *pl*, *hl*, *ql*, *fa*, *fb*, *qr*, *hr*, and *pr* indices, which stand for profile left (head turned about 90 degree left), half left (head turned about 67.5 degrees left), quarter left (head turned about 22.5 degrees left), regular frontal image, alternative frontal image, taken shortly after the corresponding *fa* image, quarter right (head turned about 22.5 degrees right), half right (head turned about 67.5 degrees right), and profile right (head turned about 90 degree right), respectively. All images have size of 512×768 pixels and are in PPM (Portable Pixel Map) format. Besides, the selected images are with and without glasses. Fig. 3 shows some images from FERET database.

We utilized MATLAB® software to evaluate the performance of our method. At first, we cropped the face area from input image manually. Then, all images are resized to 120×90 pixels. The *fa* and *fb* indices are considered as one class. So, we have 7 pose classes. From each pose of database we randomly select 150 images (totally 1050 images). From each pose, we choose 10 images as train images and the other remained images are considered as test images.

B. Feature Classification

Among the 6 sub-bands obtained from contourlet transform as shown in Fig. 4, we use those sub-bands that labeled as *b*, *c*, *e*, and *f* and are 30×23 pixels. Also, sub-bands *d* and *g* are 60×90 pixels. We consider the value of each pixel as one feature. Therefore, the dimension of feature space is 30×23×4 which is equal to 2760. This dimensionality is high and classifying such large number of features requires high complexity. Also, all of these features are not proper and many

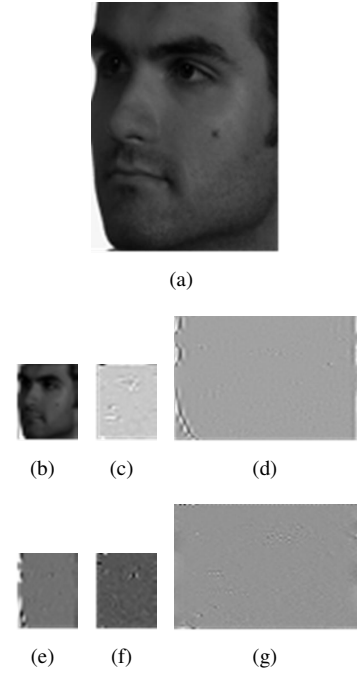


Fig. 4. (a) a typical gray-scale image, (b)-(g) sub-bands obtained after contourlet transform.

of them are redundant. As mentioned previously, in order to remove the redundancy from feature space, PCA and LDA are used. For this purpose, the normalized cumulative summation of eigenvalue corresponding to *i*th feature is calculated as follows:

$$NCsEv(i) = \frac{\sum_{n=1}^i Ev(n)}{\sum_{n=1}^M Ev(n)} \quad i = 1, \dots, M \quad (1)$$

where $Ev(x)$ is the eigenvalue of the *n*th feature, and *M* is the total number of features. $NCsEv$ is shown Fig. 5. According to $NCsEv(i)$ of features, we keep 3 features and remove others. In order to have a better representation, we only show $NCsEv$ for features 1 to 31. For features from 32 to 2760, $NCsEv$ is equal to 1. The scatter plots of the features obtained by PCA and LDA are shown in Fig. 6. Scatter plot shows the values of features for different classes. It is clear that the proposed method has good ability in separating different pose classes.

In Table I and Table II, the confusion matrices of proposed method after applying k-NN and minimum distance classifiers are presented, respectively. It is clear that proposed method can classify different poses truly (i.e.: 100% classification accuracy).

Islam *et al.* [16] used bilateral filtering and wavelet transform in pre-processing step. They used isometric projection based subspace learning for extraction of discriminant feature vectors. They chose 5000 images from FERET database, but used five different poses, as *pl*, *ql*, *fa*, *qr*, and *ql* face representations. PCA, LDA, and isometric projection (IsoP) are used separately for dimensionality reduction. For classification of feature vectors, they used the k-NN classifier. They achieved 90.338%, 95.22%, and 97.07% recognition rates for PCA,

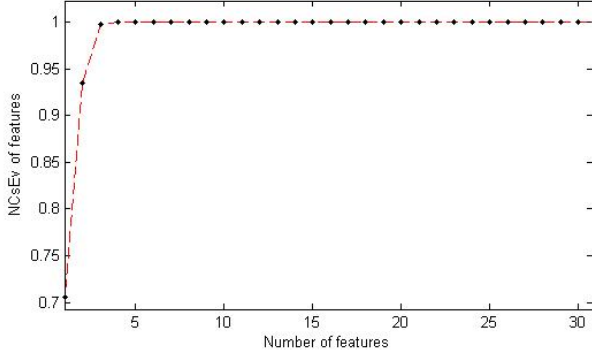


Fig. 5. Normalized cumulative summation of eigenvalues versus number of features.

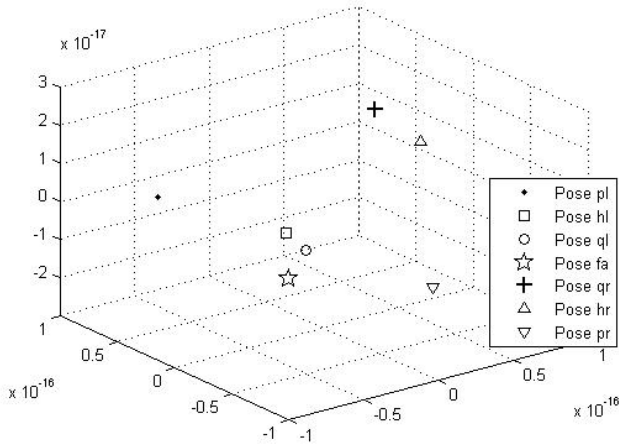


Fig. 6. Scatter plot of features obtained by PCA and LDA.

LDA, and IsoP approaches, respectively. Comparison between the results of our method and the best results of [16] is presented in Table III. From Table III, it is obvious that the proposed method has better accuracy than [16] even in the case that we use 7 classes.

V. CONCLUSION

In this paper we proposed a novel method for head pose estimation with less amount of features. In feature extraction step, we used contourlet transform. Then, we used PCA and LDA approaches for reduction of the dimensionality of the feature space. To evaluate the proposed algorithm, 1050 images from seven different poses of FERET database were selected. k-NN and minimum distance classifiers were used to classify the different poses. Experimental results demonstrate that the proposed algorithm achieve 100% classification rate on images selected from FERET database.

REFERENCES

[1] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607–626, Apr. 2009.

TABLE I
CONFUSION MATRIX FOR K-NN CLASSIFIER.

		Output Class						
		<i>pl</i>	<i>hl</i>	<i>ql</i>	<i>fa</i>	<i>qr</i>	<i>hr</i>	<i>pr</i>
Target Class	<i>pl</i>	140	0	0	0	0	0	0
	<i>hl</i>	0	140	0	0	0	0	0
	<i>ql</i>	0	0	140	0	0	0	0
	<i>fa</i>	0	0	0	140	0	0	0
	<i>qr</i>	0	0	0	0	140	0	0
	<i>hr</i>	0	0	0	0	0	140	0
	<i>pr</i>	0	0	0	0	0	0	140

TABLE II
CONFUSION MATRIX FOR MINIMUM DISTANCE CLASSIFIER.

		Output Class						
		<i>pl</i>	<i>hl</i>	<i>ql</i>	<i>fa</i>	<i>qr</i>	<i>hr</i>	<i>pr</i>
Target Class	<i>pl</i>	140	0	0	0	0	0	0
	<i>hl</i>	0	140	0	0	0	0	0
	<i>ql</i>	0	0	140	0	0	0	0
	<i>fa</i>	0	0	0	140	0	0	0
	<i>qr</i>	0	0	0	0	140	0	0
	<i>hr</i>	0	0	0	0	0	140	0
	<i>pr</i>	0	0	0	0	0	0	140

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS.

Method	Proposed method with 7 classes	Proposed method with 4 classes	Method of [16] with 5 classes
Accuracy	100%	100%	97.07%

[2] S. Ba and J. Odobez, "A probabilistic framework for joint head tracking and pose estimation," in *Proc. 17th Int'l. Conf. Pattern Recognition, 2004. ICPR 2004*, vol. 4, Aug. 2004, pp. 264–267.

[3] M. Osadchy, Y. LeCun, R., and Miller, "Synergistic face detection and pose estimation with energy-based models," *Journal of Machine Learning Research*, vol. 8, pp. 1197–1215, May 2007.

[4] J. G. Wang and E. Sung, "Em enhancement of 3d head pose estimated by point at infinity," *IEEE Trans. Image and Vision Computing*, vol. 25, no. 12, pp. 1864–1874, 2007.

[5] S. Yan, H. Wang, Y. Fu, J. Yan, X. Tang, and T. Huang, "Synchronized submanifold embedding for person-independent pose estimation and beyond," *IEEE Trans. Image Processing*, vol. 18, no. 1, pp. 202–210, Jan. 2009.

[6] V. Balasubramanian, J. Ye, and S. Panchanathan, "Biased manifold embedding: A framework for person-independent head pose estimation," in *IEEE Conf. Computer Vision and Pattern Recognition 2007. CVPR '07.*, Jun. 2007, pp. 1–7.

[7] X. Liu, H. Lu, and H. Luo, "Smooth multi-manifold embedding for robust identity-independent head pose estimation," *Computer Analysis of Images and Patterns (CAIP)*, pp. 66–73, 2009.

[8] M. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[9] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Communications*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[10] M. Do and M. Vetterli, "Pyramidal directional filter banks and curvelets," in *Proc. Int'l. Conf. Image Processing, 2001.*, vol. 3, 2001, pp. 158–161.

[11] R. Bamberger and M. Smith, "A filter bank for the directional decomposition of images: theory and design," *IEEE Trans. Signal Processing*, vol. 40, no. 4, pp. 882–893, Apr. 1992.

[12] M. Do, "Contourlet transform toolbox," <http://www.ifp.illinois.edu/~minhdo/>.

- [13] D.-Y. Po and M. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Processing*, vol. 15, no. 6, pp. 1610–1620, Jun. 2006.
- [14] S.-M. Phoong, C. Kim, P. Vaidyanathan, and R. Ansari, "A new class of two-channel biorthogonal filter banks and wavelet bases," *IEEE Trans. Signal Processing*, vol. 43, no. 3, pp. 649–665, Mar. 1995.
- [15] database available at <http://face.nist.gov/colorferet/>, "FERET."
- [16] Ihtesham-Ul-Islam, A. Khan, and I. Kim, "Effective face pose classification method," in *Proc. 2nd Int'l. Conf. Computer, Control and Communication, 2009. IC4 2009.*, Feb. 2009, pp. 1–6.